

# GENERATING INFINITE RANDOM GRAPHS

CSABA BIRÓ AND UDAYAN B. DARJI

**ABSTRACT.** Consider the following process: given a sequence of nonnegative integers  $\{d_n\}_{n=0}^\infty$  with the property that  $d_i \leq i$  (in particular  $d_0 = 0$ ), construct a random graph on countably infinitely many vertices  $v_0, v_1, \dots$ , with the following process: each new vertex  $i$  is connected to a  $d_i$ -subset of the vertices  $\{v_0, \dots, v_{i-1}\}$ , where the subset is chosen uniformly at random. We study the resulting probability space.

## 1. INTRODUCTION

Recall that the Rado graph is the unique infinite graph on countably many vertices with the property that any ordered pair of finite sets of vertices  $(U, V)$ , there is a vertex  $x$  such that  $x$  is adjacent to every vertex in  $U$ , and no vertex in  $V$ . It is customary to call the vertex  $x$  the *witness* of  $(U, V)$ . The Rado graph is universal in the sense that it contains every countable graph as an induced subgraph.

Let  $p$  be a real number with  $0 < p < 1$ . A classical result by Erdős and Rényi [2] states the following. If we construct a random graph on countably infinitely many vertices by adding an edge between any pair of vertices with probability  $p$ , then with probability 1 (we will use the standard terminology *almost surely*, and abbreviate it with *a.s.*), the resulting graph will be isomorphic to the Rado graph.

Though fascinating, the Erdős–Rényi result leaves no question open, so we are motivated to find a different random graph model that results in potentially different probability spaces. Since every vertex of the Rado graph a.s. has infinite degree, we start by adding only finitely many incident vertices to each vertex, determined by a given sequence.

More rigorously, suppose a sequence of integers  $\{d_i\}_{i=0}^\infty$  is given, with the property  $0 \leq d_i \leq i$  for all  $i$ . Let  $V = \{v_0, v_1, \dots\}$  be a set of vertices. Then for  $i = 0, 1, \dots$ , in round  $i$ , choose a subset of  $A$  of  $\{v_0, \dots, v_{i-1}\}$  uniformly at random. Then add edges  $v_i u$  for all  $u \in A$ . The result is a random graph on countably many vertices. We strive to understand the resulting probability space, in particular we would like to determine the atoms (graphs with positive probability), and cases when there is only one atom with probability 1. In this latter case, we say that the probability space is *concentrated*.

**1.1. Non-concentrated spaces.** It is perhaps not completely naive to think that something similar happens here as in the Erdős–Rényi model. In this subsection we demonstrate that is far from being correct. Thereby we will show examples of non-concentrated spaces.

We will use the term  $\omega$ -tree for the unique countably infinite tree in which every vertex is of infinite degree.

The following proposition is actually about a very simple example of *concentration*, but we will use it as a tool to show non-concentration in some other cases.

**Proposition 1.1.** *The sequence  $0, 1, 1, 1, \dots$  a.s. generates the  $\omega$ -tree.*

*Proof.* We will prove a more general statement later, see Theorem 3.15.  $\square$

**Corollary 1.2.** *A finite sequence and then  $1, 1, 1, \dots$  a.s. generates the following random graph: Generate the random graph of the finite sequence, and attach an  $\omega$ -tree to each vertex.*

The corollary above shows that it is easy to construct a sequence whose associated probability space is not concentrated. E.g.  $0, 1, 2, 1, 2, 1, 1, 1, \dots$ . However these examples are very special in the sense that they are eventually all 0's and 1's, so after that point no more cycles are generated any more. Nevertheless, we have the following proposition.

**Proposition 1.3.** *There exists a sequence, not eventually 0's and 1's with non-concentrated probability space.*

*Proof.* We will construct a sequence consisting mostly of 1's, but infinitely many 2's inserted. The sequence starts with  $0, 1, 1, 2$ . We set  $p_0 = 2/3$ , and we note that  $p_0$  is the probability that the first 4 vertices span a triangle. Then let  $k$  be the least integer such that  $k/\binom{k}{2} < 3/4 - p_0$ . Set  $d_4 = \dots = d_{k-1} = 1$ , and  $d_k = 2$ . Note, that the probability that a triangle is generated from  $v_k$  is  $p_1 := k/\binom{k}{2}$ . In general, after the  $l$ th 2 in the sequence, let  $k$  be a sufficiently large integer for which  $d_k$  is not yet defined and

$$\frac{k+l-1}{\binom{k}{2}} < \frac{3}{4} - \sum_{i=0}^{l-1} p_i.$$

Set  $d_k = 2$  and set all the elements before  $d_k$  that are not yet defined to be 1. Note that the probability that a triangle is generated at  $v_k$  equals  $p_l = \frac{k+l-1}{\binom{k}{2}}$ . Let  $X$  be the random variable that denotes the number of triangles eventually generated in  $G$ . Due to the linearity of expectation,

$$\mu = E[X] = \sum_{i=0}^{\infty} p_i.$$

Clearly, from the definition of the sequence  $2/3 \leq \mu \leq 3/4$ . That means that

$$\Pr[X = 0] > 0 \quad \text{and} \quad \Pr[X > 0] \geq \Pr[X = 1] = 2/3 > 0.$$

The sets  $[X = 0]$  and  $[X > 0]$  partition the probability space, and neither of them are of measure 0, so the space can not be concentrated.  $\square$

## 2. THE RADO GRAPH

We start with a simple technical lemma. We will use the standard notation  $n_{(k)} = n(n-1)\dots(n-k+1)$ .

**Lemma 2.1.** *Fix a  $k$  positive integer. The infinite series*

$$\sum_n \left(\frac{d_n}{n}\right)^k \left(\frac{n-d_n}{n}\right)^k \quad \text{and} \quad \sum_n \frac{(d_n)_{(k)}(n-d_n)_{(k)}}{(n)_{(2k)}}$$

*either both converge or both diverge.*

*Proof.* If  $k = 1$  then the statement is trivial. If  $k \geq 2$ , then partition the terms into 3 parts:  $A = \{i : d_i < k\}$ ,  $B = \{i : n - d_i < k\}$ , and  $C = \mathbb{N} \setminus (A \cup B)$ . It is clear that over the terms indexed by  $A$  and  $B$ , both series converge, so the behavior is decided by the terms over  $C$ . For those we use a generalized limit comparison test, and show that the lim inf and lim sup of the ratio of the terms are positive and finite.  $\square$

The following definition provides a notion about how “similar” a graph is to the Rado graph.

**Definition 2.2.** *Let  $G$  be a graph. For a  $k$  nonnegative integer, we say  $G$  is  $k$ -Rado, if for every  $(A, B)$  pair of disjoint finite pair of set of vertices such that  $|A| \leq k$  and  $|B| \leq k$  there are infinitely many witnesses.*

*The number  $\text{rado}(G) = \max\{k : G \text{ is } k\text{-Rado}\}$  is the radocity of  $G$ .*

Clearly every graph is 0-Rado, and if a graph is  $k$ -Rado, it is also  $k'$ -Rado for all  $k' < k$ . Also,  $G$  is isomorphic to the Rado graph if and only if  $\text{rado}(G) = \infty$ .

**Theorem 2.3.** *Let*

$$k = \max \left\{ t \in \mathbb{N}^+ : \sum_{n=1}^{\infty} \left( \frac{d_n}{n} \right)^t \left( \frac{n - d_n}{n} \right)^t = \infty \right\}$$

$$= \max \left\{ t \in \mathbb{N}^+ : \sum_{n=1}^{\infty} \frac{(d_n)_{(k)}(n - d_n)_{(k)}}{(n)_{(2k)}} = \infty \right\}.$$

*Then the process a.s. generates a graph of radocity  $k$ . (We take the maximum of an empty set to be 0 here.)*

*Proof.* Note that the two expressions that define  $k$  in the statement are equivalent due to Lemma 2.1. We will use this fact to go back and forth between the two definitions at our convenience.

First we prove that the graph generated is almost surely  $k$ -Rado.

Let  $A, B$  be two finite disjoint vertex sets with  $|A| = |B| = k \geq 1$ , and let  $N$  be a positive integer. It is sufficient to show that the pair  $(A, B)$  has a witness with probability 1 among the vertices  $v_N, v_{N+1}, \dots$

For a given vertex  $v_n$ , let  $p_n$  be the probability that  $v_n$  is a witness for  $(A, B)$ . Now pick a vertex  $v_n$  such that  $n \geq N$ , and  $n > \max\{i : v_i \in A \cup B\}$ . Then

$$p_n = \frac{\binom{n-2k}{d_n-k}}{\binom{n}{d_n}} = \frac{(n-2k)!}{(n-k-d_n)!(d_n-k)!} \cdot \frac{d_n!(n-d_n)!}{n!} = \frac{(d_n)_{(k)}(n-d_n)_{(k)}}{(n)_{(2k)}}$$

so  $\sum_{n=N}^{\infty} p_n$  diverges. Therefore  $\prod_{n=N}^{\infty} (1 - p_n) = 0$ , which is the probability that the pair  $(A, B)$  has no witness beyond (including)  $v_N$ .

It remains to be proven that if  $k < \infty$ , then the graph is a.s. not  $k+1$ -Rado. Suppose for a contradiction that this is not true: that in fact the probability that the graph is not  $k+1$ -Rado has probability  $p < 1$ . Now fix any pair of disjoint sets of vertices  $A, B$  such that  $|A| = |B| = k+1$ . Let  $q_N$  be the probability that  $(A, B)$  has no witness beyond (including)  $v_N$ . Since this event implies that the graph is not  $k+1$ -Rado, it must be that

$$(1) \quad q_N \leq p \text{ for all } N.$$

On the other hand, similarly as above, the probability that a given vertex  $v_n$  is a witness for  $(A, B)$  (if  $n$  is large enough) is

$$p_n = \frac{(d_n)_{(k+1)}(n - d_n)_{(k+1)}}{(n)_{(2(k+1))}}.$$

This time, we know that  $\sum p_n < \infty$ , so  $\prod(1 - p_n) > 0$ . Hence there exists  $N$  such that

$$q_N = \prod_{i=N}^{\infty} (1 - p_i) > p.$$

But this contradicts (1).  $\square$

**Corollary 2.4.** *Let  $a_n = \min\{\frac{d_n}{n}, \frac{n-d_n}{n}\}$ .*

- (i) *If  $\sum_{n=1}^{\infty} a_n^k$  diverges for all  $k$  positive integer, then the process almost surely generates the Rado graph.*
- (ii) *If there is a  $k$  positive integer for which  $\sum_{n=1}^{\infty} a_n^k$  converges, then the process almost surely does not generate the Rado graph.*

*Proof.* Suppose that  $\sum_{n=1}^{\infty} a_n^k$  diverges for all  $k$ . Since

$$\sum_{n=1}^{\infty} \left(\frac{d_n}{n}\right)^k \left(\frac{n-d_n}{n}\right)^k \geq \sum_{n=1}^{\infty} a_n^{2k},$$

we get that

$$\sum \left(\frac{d_n}{n}\right)^k \left(\frac{n-d_n}{n}\right)^k$$

diverges for all  $k$ , and therefore we get a.s.  $\text{rado}(G) = \infty$ .

Now suppose there is a positive integer  $k$  for which  $\sum_{n=1}^{\infty} a_n^k$  converges. If  $n_0$  is large enough, then

$$\begin{aligned} \sum_{n=n_0}^{\infty} a_n^k \cdot 2^{2k} &\geq \sum_{n=n_0}^{\infty} \left( \prod_{i=0}^{k-1} \frac{d_n - i}{n} \cdot \frac{n - d_n - i}{n} \prod_{i=0}^{2k-1} \frac{n}{n - i} \right) \\ &= \sum_{n=n_0}^{\infty} \left( \prod_{i=0}^{k-1} \frac{d_n - i}{n} \cdot \frac{n}{n - i} \prod_{i=0}^{k-1} \frac{n - d_n - i}{n} \cdot \frac{n}{n - k - i} \right) \\ &= \sum_{n=n_0}^{\infty} \frac{(d_n)_{(k)}(n - d_n)_{(k)}}{(n)_{(2k)}}, \end{aligned}$$

where the last sum converges. Thus the graph a.s. has finite radocity.  $\square$

**Corollary 2.5.** *Let  $a_n$  be as in Corollary 2.4. If  $\limsup a_n > 0$ , then the process a.s. generates the Rado graph.*

The *double random process* is when we even chose the sequence in random, choosing  $d_i$  with some distribution from the interval  $[0, i]$ . The following corollary states that in some sense, almost all double random processes will result in the Rado graph.

**Corollary 2.6.** *If there exist  $\epsilon > 0$ ,  $p_0 > 0$ , and  $M$  integer such that for  $n > M$ ,  $\Pr[en \leq d_n \leq (1 - \epsilon)n] \geq p_0$ , then the double random process a.s. generates the Rado graph.*

*Proof.* It is easy to see that Corollary 2.5 is satisfied.  $\square$

## 3. ZERO-ONE SEQUENCES

In the later parts of this section we will focus on sequences with  $d_i \in \{0, 1\}$ . However, we won't need this condition at the beginning, so we prove a few theorems in the general setting. We will use these later to essentially characterize the probability spaces in the zero-one case.

**3.1. Notations.** For the rest of the section it will be convenient to introduce many notations to denote certain tuples of indices and sums and products. First we introduce notations on products and tuples.

We let  $\mathbb{N}^{<\mathbb{N}}$  denote set of all finite strings of  $\mathbb{N}$  including the empty string. We define  $f : \mathbb{N}^{<\mathbb{N}} \rightarrow [0, \infty)$  by

$$f(\sigma) = \prod_{i=1}^n \frac{d_{\sigma_i}}{\sigma_i},$$

where  $\sigma = (\sigma_1, \dots, \sigma_n)$ . (By convention,  $f(\sigma) = 1$  when  $\sigma$  is the empty string.) Suppose  $i, j, l \geq 1$  with  $i \leq j$ . Then,

$$A_i^l = \{(\sigma_1, \dots, \sigma_l) \in \mathbb{N}^l \mid \min\{\sigma_1, \dots, \sigma_l\} = i\}$$

$$B_i^l = \{\sigma \in A_i^l \mid \sigma \text{ is increasing}\}$$

$$B_{i,j}^l = \{(\sigma_1, \dots, \sigma_l) \in B_i^l \mid \max\{\sigma_1, \dots, \sigma_l\} = j\}$$

$$C_i^l = \{\sigma \in A_i^l \mid \sigma \text{ is injective}\}$$

$$D_i^l = \{i, i+1, \dots\}^l = \{(\sigma_1, \dots, \sigma_l) \in \mathbb{N}^l \mid \min\{\sigma_1, \dots, \sigma_l\} \geq i\}$$

The following notations are about sums and series.

$$s_{m,n} = \sum_{i=m}^n d_i \quad s_n = s_{0,n} = \sum_{i=0}^n d_i \quad t_{n,m} = \sum_{i=n}^m \frac{d_i}{i} \quad t_n = t_{n,\infty} = \sum_{i=n}^{\infty} \frac{d_i}{i}$$

**3.2. General findings.** We begin with a simple proposition on binomial coefficients.

**Proposition 3.1.** *Let  $n, d, m \geq 0$  integers with  $\frac{m}{n-d} \leq 1$ . Then*

$$\left(1 - \frac{m}{n-d}\right)^d \leq \frac{\binom{n-m}{d}}{\binom{n}{d}} \leq \left(1 - \frac{m}{n}\right)^d.$$

*Proof.* We note that

$$\frac{\binom{n-m}{d}}{\binom{n}{d}} = \frac{(n-m)_{(d)}}{(n)_{(d)}} = \prod_{i=0}^{d-1} \left(1 - \frac{m}{n-i}\right).$$

Then bound the product by replacing all factors with the largest factor, and then with the smallest factor to obtain the desired inequality.  $\square$

**Lemma 3.2.** *Let  $V = \{v_{i_1}, v_{i_2}, \dots, v_{i_k}\}$  be a finite set of vertices where  $i_1 < i_2 < \dots < i_k$ .*

- i) If  $\sum d_i/i = \infty$  then for all  $N > i_k$ , a.s.  $V$  has a neighbor beyond  $v_N$ .*
- ii) If  $\sum d_i/i < \infty$  then there exists  $M$  such that with positive probability  $V$  has no neighbor beyond  $v_M$ ; furthermore, for all  $\epsilon > 0$  there exists an  $M' \geq M$  such that  $\Pr(V \text{ has a neighbor beyond } v_{M'}) < \epsilon$ .*

*Proof.* Let  $E_l$  be the event that  $V$  has no neighbor beyond  $v_l$ . We will estimate the probability of  $E_l$ . For any  $i > i_k$ , we have  $\Pr(v_i \not\sim V) = \binom{i-k}{d_i} / \binom{i}{d_i}$ , so  $\Pr(E_l) = \prod_{i=l}^{\infty} \binom{i-k}{d_i} / \binom{i}{d_i}$ .

Using Proposition 3.1, we have

$$\Pr(E_l) \leq \prod_{i=l}^{\infty} \left(1 - \frac{k}{i}\right)^{d_i}.$$

Now if  $\sum d_i/i = \infty$ , then  $\sum kd_i/i = \infty$ , which implies the product on the right hand side is zero. Thus, we have that  $\Pr(E_l) = 0$  for all  $l > i_k$ , which proves the first part.

If  $\sum d_i/i < \infty$ , then there exists an  $M$  such that for all  $i \geq M$ ,  $d_i/i \leq 1/2$ , and  $k/(i - d_i) \leq 1$ . Then we may use the other part of Proposition 3.1 to get

$$(2) \quad \Pr(E_M) \geq \prod_{i=M}^{\infty} \left(1 - \frac{k}{i - d_i}\right)^{d_i}.$$

Also, in this case,

$$\sum_{i=M}^{\infty} \frac{kd_i}{i - d_i} = \sum_{i=M}^{\infty} \frac{k}{1 - d_i/i} \cdot \frac{d_i}{i} \leq \sum_{i=M}^{\infty} 2k \cdot \frac{d_i}{i} < \infty,$$

so  $\Pr(E_M) > 0$ .

The last statement follows from the fact that the right hand side in (2) is positive, therefore its tail end converges to 1, so for all  $\epsilon > 0$  there exists  $M'$  for which  $\Pr(E_{M'}) > 1 - \epsilon$ .  $\square$

**Theorem 3.3.** *The following statements hold.*

- i) *If  $\sum d_i/i = \infty$ , then the process a.s. generates a graph in which each vertex is of infinite degree.*
- ii) *If  $\sum d_i/i < \infty$ , then the process a.s. generates a graph in which each vertex is of finite degree.*

*Proof.* Both parts follow from Lemma 3.2. Consider a vertex  $v_k$  and an integer  $N > k$ . Apply Lemma 3.2 for  $V = \{v_k\}$ , and  $N$ . In case i), we conclude that a.s.  $v_k$  has a neighbor beyond  $v_N$ . This being true for arbitrary  $N > k$ , we conclude that a.s.  $v_k$  has infinitely many neighbors.

In case ii), the lemma provides that the probability that  $v_k$  is of infinite degree is less than  $\epsilon$  for all  $\epsilon > 0$ , and therefore that probability is 0.  $\square$

The following is simple interesting fact that prove very useful later.

**Theorem 3.4.** *Suppose  $\sum d_i/i < \infty$ . Then the process a.s. generates a graph with no ray.*

*Proof.* The hypothesis implies that  $t_n \rightarrow 0$ , so there exists  $N$  positive integer such that for all  $i \geq N$ ,  $t_n < 1$ . Fix  $i \geq N$ . For  $k \geq 2$ , let  $E_k$  denote the event that the graph will contain a path of length  $k$  that starts at the vertex  $v_i$  and every vertex of the path is beyond  $v_i$ .

$$\begin{aligned}
\Pr(E_k) &\leq \sum_{\sigma \in D_{i+1}^k} \Pr(v_{\sigma_{l+1}} \sim v_{\sigma_l} \text{ for all } 0 \leq l < k) \\
&= \sum_{\sigma \in D_{i+1}^k} \frac{d_{\sigma_1}}{\sigma_1} \dots \frac{d_{\sigma_k}}{\sigma_k} = \sum_{\sigma \in D_{i+1}^k} f(\sigma) = t_{i+1}^k
\end{aligned}$$

Hence  $\lim_{k \rightarrow \infty} \Pr(E_k) = 0$ , which implies the statement.  $\square$

Recall that the bipartite graphs  $K_{1,l}$  for  $l = 1, 2, \dots$  are called *stars*. To emphasize the size of the star,  $K_{1,l}$  will often be called an  $l$ -star. We say a vertex in a graph is *in a star*, respectively *in an  $l$ -star*, if the connected component of the vertex is a star, respectively an  $l$ -star.

**Lemma 3.5.** *Suppose  $\sum s_i d_i / i < \infty$ . Then the process a.s. generates a graph  $G$  which has the property that there exists an  $N_1 = N_1(G)$  such that for all  $n \geq N_1$  with  $d_n > 0$ , the vertex  $v_n$  will attach back to vertices with current degree 0.*

*Proof.* Note that  $\sum s_i d_i / i < \infty$  implies  $s_n d_n / n \rightarrow 0$  as  $n \rightarrow \infty$ , so there exists  $N$  such that for all  $n > N$ ,  $s_n d_n / n < 1/3$ . Consider such an  $n$ . Below we will compute the probability that at stage  $n$ , the vertex  $v_n$  attaches only to the vertices that are currently of degree 0, i.e., singletons.

Observe that during the process, for every  $i$  with  $d_i = 0$  one singleton is created, and if  $d_i > 0$ , then at most  $d_i$  singletons are destroyed. One can view this as creating a singleton and then destroying at most  $d_i + 1 \leq 2d_i$  in all cases. So at step  $i$ , the number of singletons is at least  $i - \sum_{j=0}^i 2d_j = i - 2s_i$ , and then, by Proposition 3.1 and the fact that  $d_n / n < 1/3$ , we obtain that the probability that  $v_n$  attaches to all singletons is at least

$$\begin{aligned}
\frac{\binom{n-2s_n}{d_n}}{\binom{n}{d_n}} &\geq \left(1 - \frac{2s_n}{n-d_n}\right)^{d_n} \geq \left(1 - \frac{2s_n/n}{1-d_n/n}\right)^{d_n} \\
&\geq \left(1 - \frac{2s_n/n}{2/3}\right)^{d_n} = \left(1 - \frac{3s_n}{n}\right)^{d_n}.
\end{aligned}$$

Hence the probability that this happens to all vertices beyond  $N$  is at least

$$\prod_{n=N}^{\infty} \left(1 - \frac{3s_n}{n}\right)^{d_n}.$$

The last product is positive as  $\sum s_i d_i / i < \infty$ . Hence, we have that for all  $\epsilon > 0$  there exists  $M$ , such that  $\prod_{n=M}^{\infty} \left(1 - \frac{3s_n d_n}{n}\right) > 1 - \epsilon$ . That means that with probability greater than  $1 - \epsilon$ , every vertex beyond  $M$  attaches to singletons, completing the proof.  $\square$

**Theorem 3.6.** *Suppose  $\sum s_i d_i / i < \infty$ . Then the process a.s. generates a graph  $G$  for which there is  $N = N(G)$  such that for all  $n > N$ ,  $v_n$  is in a star. Moreover, if, in addition,  $d_n > 0$ , then  $v_n$  is in  $d_n$ -star.*

*Proof.* By Lemma 3.5, the process a.s. generates a graph  $G$  for which there is  $N = N(G)$  such that for all  $n > N$ , at stage  $n$ , either  $d_n = 0$  or  $v_n$  attaches to  $d_n$  many current degree 0 vertices which precede  $v_n$ . Note that  $v_k$ ,  $k > n$ , leaves

untouched the star generated by  $v_n$ . Therefore, we obtain that  $v_n$  is in a  $d_n$ -star. If  $n > N$  and  $d_n = 0$ , then the component of  $v_n$  in  $G$  is either a singleton or a star generated by some  $v_m$ ,  $m > n$ .  $\square$

**The sparse case.** We are ready to consider sequences with  $0 \leq d_i \leq 1$ . We will also assume that there are infinitely many 1's in the sequence, for otherwise we really have a finite sequence and an essentially finite graph (plus isolated vertices), and we get a problem of a very different flavor. It is clear that the number of connected components of the generated graph is equal to the number of 0's in the sequence, and each component is a tree.

What we proved in the first part of this section essentially gives us the behavior of the probability space in the dense case, when  $\sum d_i/i = \infty$ , and the very sparse case, when  $\sum s_i d_i/i < \infty$ . We will summarize these findings (and much more) in Theorem 3.15. This subsection will entirely be devoted to the sparse case. Accordingly, throughout the subsection we will assume that  $\sum d_i/i < \infty$ . Our findings *will* apply for the very sparse case, giving an alternative proof of the characterization of the space in that case. But note, that the findings at the beginning of this section apply for the very sparse case *in the general setting* (not only zero-one), so those theorems still have their importance.

**Proposition 3.7.**  $\lim_{i \rightarrow \infty} s_i/i = 0$ .

*Proof.* Let  $\epsilon > 0$  small. Then,

$$\frac{s_n}{n} = \frac{\sum_{i=0}^n d_i}{n} \leq \frac{\sum_{i=0}^{\lfloor \epsilon n/2 \rfloor} d_i}{n} + \sum_{i=\lfloor \epsilon n/2 \rfloor+1}^n \frac{d_i}{i} \leq \frac{\epsilon}{2} + \sum_{i=\lfloor \epsilon n/2 \rfloor+1}^{\infty} \frac{d_i}{i}.$$

As  $\sum d_i/i < \infty$ , the second term converges to zero as  $n \rightarrow \infty$ , so eventually it will be less than  $\epsilon/2$ .  $\square$

Let  $a(k)_i$  denote the expected number of trees of size  $k$  spanned by the vertex set  $\{v_0, \dots, v_i\}$ . We will prove a sequence of technical lemmas about the sequences  $a(k)$ .

**Proposition 3.8.** For  $i \geq 1$  and  $k \geq 2$ ,

$$a(k)_i = a(k)_{i-1} + \frac{(k-1)a(k-1)_{i-1}}{i}d_i - \frac{ka(k)_{i-1}}{i}d_i.$$

*Proof.* Suppose  $d_i = 1$ . Consider the process just before we add the edge from  $v_i$ . Let  $p_+$  be the probability that we increase the number of trees of size  $k$ , and let  $p_-$  be the probability that we decrease that number. In either case, the change is  $\pm 1$ . Let  $p_0 = 1 - (p_+ - p_-)$ . On one hand,  $p_+ = \frac{(k-1)a(k-1)_{i-1}}{i}$ , and  $p_- = \frac{ka(k)_{i-1}}{i}$ . On the other hand

$$a(k)_i = p_+(a(k)_{i-1} + 1) + p_-(a(k)_{i-1} - 1) + p_0 a(k)_{i-1} = a(k)_{i-1} + (p_+ - p_-).$$

In the other case, if  $d_i = 0$ , then  $a(k)_i = a(k)_{i-1}$ .  $\square$

**Lemma 3.9.** Let  $k \geq 2$ . Then there exist positive constants  $C_1, C_2$  such that for all  $i \geq 1$

$$C_1 \sum_{j=k}^i \frac{a(k-1)_{j-1}}{j} d_j \leq a(k)_i \leq C_2 \sum_{j=1}^i \frac{a(k-1)_{j-1}}{j} d_j.$$



*Proof.* The upper bound is a straightforward consequence of Proposition 3.8. Indeed,  $a(k)_i \leq a(k)_{i-1} + \frac{ka(k-1)_{i-1}}{i}d_i$ , so  $a(k)_i \leq k \sum_{j=1}^i \frac{a(k-1)_{j-1}}{j}d_j$ .

For the lower bound, notice that

$$a(k)_i \geq a(k)_{i-1} \left(1 - \frac{kd_i}{i}\right) + \frac{a(k-1)_{i-1}}{i}d_i.$$

This implies

$$\begin{aligned} a(k)_i &\geq \sum_{j=1}^i \left[ \frac{a(k-1)_{j-1}}{j}d_j \prod_{l=j+1}^i \left(1 - \frac{kd_l}{l}\right) \right] \\ &\geq \sum_{j=k}^i \left[ \frac{a(k-1)_{j-1}}{j}d_j \prod_{l=j+1}^i \left(1 - \frac{kd_l}{l}\right) \right] \geq \sum_{j=k}^i \left[ \frac{a(k-1)_{j-1}}{j}d_j \prod_{l=k+1}^{\infty} \left(1 - \frac{kd_l}{l}\right) \right] \\ &\geq Q \sum_{j=k}^i \frac{a(k-1)_{j-1}}{j}d_j, \end{aligned}$$

where  $Q = \prod_{l=k+1}^{\infty} \left(1 - \frac{kd_l}{l}\right)$ . Note that the second inequality is correct, because for  $j = 1, \dots, k-2$ , we have  $a(k-1)_{j-1} = 0$  (so the omitted terms are zero), and for  $j = k-1$  the omitted term is nonnegative. Also note that  $\sum d_i/i < \infty$  implies  $Q > 0$ .  $\square$

**Lemma 3.10.** *Let  $k \geq 2$ . Then there exists positive constants  $K$  such that for all  $i$ ,*

$$a(k)_i \leq K \sum_{j=1}^i d_j t_{j+1}^{k-2}.$$

where  $t_j = \sum_{l=j}^{\infty} \frac{d_l}{l}$ .

*Proof.* We proceed by induction on  $k$ . Let  $k = 2$ . There exists  $C_2$  such that  $a(2)_i \leq C_2 \sum_{j=1}^i \frac{a(1)_{j-1}}{j}d_j \leq C_2 \sum_{j=1}^i d_j$ .

Now suppose  $k \geq 3$ . There exists  $C_2$  and  $C$  positive constants such that

$$\begin{aligned} a(k)_i &\leq C_2 \sum_{j=1}^i \frac{a(k-1)_{j-1}}{j}d_j \leq C \sum_{j=2}^i \sum_{l=1}^{j-1} d_l t_{l+1}^{k-3} \frac{d_j}{j} \leq C \sum_{l=1}^{i-1} \sum_{j=l+1}^i d_l t_{l+1}^{k-3} \frac{d_j}{j} \\ &\leq C \sum_{l=1}^{i-1} d_l t_{l+1}^{k-3} \sum_{j=l+1}^i \frac{d_j}{j} \leq C \sum_{l=1}^{i-1} d_l t_{l+1}^{k-2}. \end{aligned}$$

$\square$

**Lemma 3.11.** *Let  $l \geq 1$ .*

$$\sum_{i=1}^{\infty} \sum_{\sigma \in A_i^l} f(\sigma) < \infty.$$

*Proof.* For  $l = 1$  the above statement is equivalent to the sparsity condition  $\sum d_i/i < \infty$ . For  $l > 1$ , we note that

$$\sum_{\sigma \in A_i^l} f(\sigma) \leq \sum_{j=1}^l \sum_{\substack{\sigma \in A_i^l \\ \sigma_j = i}} \frac{d_i}{i} \prod_{\substack{k=1 \\ k \neq j}}^l \frac{d_{\sigma_k}}{\sigma_k} = \sum_{j=1}^l \frac{d_i}{i} \sum_{\sigma \in D_i^{l-1}} f(\sigma) = \sum_{j=1}^l \frac{d_i}{i} t_i^{l-1} = l \frac{d_i}{i} t_i^{l-1}$$

Then,

$$\sum_{i=1}^{\infty} \sum_{\sigma \in A_i^l} f(\sigma) \leq l \sum_{i=1}^{\infty} \frac{d_i}{i} t_i^{l-1}.$$

As  $\lim_{i \rightarrow \infty} t_i = 0$ , we have that  $\{t_i\}_{i=1}^{\infty}$  is bounded and hence the desired series converges.  $\square$

**Lemma 3.12.** *Suppose that  $l \geq 1$ . Then,*

$$\sum_{i=1}^{\infty} d_i t_i^l = \infty \implies \sum_{i=1}^{\infty} s_i \sum_{\sigma \in B_i^l} f(\sigma) = \infty.$$

*Proof.* By rearranging and switching the order of summation, we have that

$$(3) \quad \sum_{i=1}^{\infty} d_i t_i^l = \sum_{i=1}^{\infty} d_i \sum_{j=i}^{\infty} \sum_{\sigma \in A_j^l} f(\sigma) = \sum_{j=1}^{\infty} \sum_{i=1}^j d_i \left( \sum_{\sigma \in A_j^l} f(\sigma) \right) = \sum_{j=1}^{\infty} s_j \sum_{\sigma \in A_j^l} f(\sigma) = \infty.$$

We next observe that if  $l = 1$ , then  $A_i^l = B_i^l$  and the proof is complete. Hence, let us assume that  $l \geq 2$ . We will next show that

$$(4) \quad \sum_{i=1}^{\infty} s_i \sum_{\sigma \in A_i^l \setminus C_i^l} f(\sigma) < \infty.$$

$$\begin{aligned}
\sum_{i=1}^{\infty} s_i \sum_{\sigma \in A_i^l \setminus C_i^l} f(\sigma) &\leq \sum_{i=1}^{\infty} s_i \sum_{1 \leq j < k \leq l} \sum_{\substack{\sigma \in A_i^l \\ \sigma_j = \sigma_k}} f(\sigma) \\
&= \sum_{i=1}^{\infty} s_i \sum_{1 \leq j < k \leq l} \left( \sum_{\substack{\sigma \in A_i^l \\ \sigma_j = \sigma_k = i}} f(\sigma) + \sum_{m=i+1}^{\infty} \sum_{\substack{\sigma \in A_i^l \\ \sigma_j = \sigma_k = m}} f(\sigma) \right) \\
&= \sum_{i=1}^{\infty} s_i \sum_{1 \leq j < k \leq l} \left( \left( \frac{d_i}{i} \right)^2 \sum_{\sigma \in D_i^{l-2}} f(\sigma) + \sum_{m=i+1}^{\infty} \sum_{\substack{1 \leq p \leq l \\ p \notin \{j, k\}}} \sum_{\substack{\sigma \in A_i^l \\ \sigma_j = \sigma_k = m \\ \sigma_p = i}} f(\sigma) \right) \\
&= \sum_{i=1}^{\infty} s_i \sum_{1 \leq j < k \leq l} \left( \left( \frac{d_i}{i} \right)^2 \sum_{\sigma \in D_i^{l-2}} f(\sigma) + \sum_{m=i+1}^{\infty} \sum_{\substack{1 \leq p \leq l \\ p \notin \{j, k\}}} \frac{d_i}{i} \frac{d_m}{m} \frac{d_m}{m} \sum_{\sigma \in D_i^{l-3}} f(\sigma) \right) \\
&\leq \sum_{i=1}^{\infty} s_i \sum_{1 \leq j < k \leq l} \left( \left( \frac{d_i}{i} \right)^2 \sum_{\sigma \in D_i^{l-2}} f(\sigma) + (l-2) \left( \frac{d_i}{i} \right)^2 \sum_{m=i+1}^{\infty} \frac{d_m}{m} \sum_{\sigma \in D_i^{l-3}} f(\sigma) \right) \\
&\leq \sum_{i=1}^{\infty} s_i \sum_{1 \leq j < k \leq l} \left( \left( \frac{d_i}{i} \right)^2 \sum_{\sigma \in D_i^{l-2}} f(\sigma) + (l-2) \left( \frac{d_i}{i} \right)^2 \sum_{\sigma \in D_i^{l-2}} f(\sigma) \right) \\
&\leq (l-1) \binom{l}{2} \sum_{i=1}^{\infty} s_i \left( \frac{d_i}{i} \right)^2 \sum_{\sigma \in D_i^{l-2}} f(\sigma) = (l-1) \binom{l}{2} \sum_{i=1}^{\infty} \frac{s_i}{i} \frac{d_i}{i} t_i^{l-2} < \infty.
\end{aligned}$$

As before, the last inequality follows as  $\{s_i/i\}_{i=1}^{\infty}$  and  $\{t_i\}_{i=1}^{\infty}$  are bounded sequences.

Putting (3) and (4) together we have that

$$\sum_{i=1}^{\infty} s_i \sum_{\sigma \in C_i^l} f(\sigma) = \infty.$$

Noting

$$\sum_{i=1}^{\infty} s_i \sum_{\sigma \in C_i^l} f(\sigma) = l! \sum_{i=1}^{\infty} s_i \sum_{\sigma \in B_i^l} f(\sigma),$$

the proof is complete.  $\square$

**Lemma 3.13.** *Let  $k \geq 2$ . Then, there exists  $C$  and  $N$  such that for all  $i \in \mathbb{N}$ , we have that*

$$\begin{aligned}
a(2)_i &\geq C s_{N,i}, \text{ and} \\
a(k)_i &\geq C \sum_{j=3}^{i+3-k} s_{N,j-1} \sum_{\sigma \in B_{j,i}^{k-2}} f(\sigma) \quad \text{for } k \geq 3.
\end{aligned}$$

*Proof.* Let  $N \geq 2$  be such that for all  $n > N$ ,  $2s_n/n \leq 1/2$ .

First consider  $k = 2$ . From Lemma 3.9 there exists a constant  $C$  such that for all  $i \geq 2$

$$a(2)_i \geq C \sum_{j=2}^i \frac{a(i)_{j-1}}{j} d_j.$$

So for all  $i \geq N$

$$\begin{aligned} a(2)_i &\geq C \sum_{j=N}^i \frac{j - 2s_{j-1}}{j} d_j \geq C \sum_{j=N}^i \frac{j - 2s_j}{j} d_j \\ &\geq C \sum_{j=N}^i \left(1 - \frac{2s_j}{j}\right) d_j \geq \frac{C}{2} \sum_{j=N}^i d_j = \frac{C}{2} s_{N,i}. \end{aligned}$$

Now assume  $k \geq 3$ . We will proceed by induction, so we consider  $k = 3$  first. There exists  $C$  such that for all  $i \geq 3$

$$a(3)_i \geq C \sum_{j=3}^i \frac{a(2)_{j-1}}{j} d_j \geq C \sum_{j=3}^i s_{N,j-1} \frac{d_j}{j} \geq C \sum_{j=3}^i s_{N,j-1} \sum_{\sigma \in B_{j,i}^1} f(\sigma).$$

Now assume that  $k \geq 4$ . There exists a  $C$  such that for all  $i$

$$\begin{aligned} a(k)_i &\geq C \sum_{j=k}^i \frac{a(k-1)_{j-1}}{j} d_j \geq C \sum_{j=k}^i \sum_{l=3}^{j-k+3} s_{N,l-1} \sum_{\sigma \in B_{l,j-1}^{k-3}} f(\sigma) \frac{d_j}{j} \\ &\geq C \sum_{l=3}^{i-k+3} s_{N,l-1} \sum_{j=l-3+k}^i \frac{d_j}{j} \sum_{\sigma \in B_{l,j-1}^{k-3}} f(\sigma) = C \sum_{l=3}^{i-k+3} s_{N,l-1} \sum_{\sigma \in B_{l,i}^{k-2}} f(\sigma). \end{aligned}$$

□

**Lemma 3.14.** *Let  $k \geq 2$ . Then,*

$$\left[ \sum_{j=1}^{\infty} d_j t_j^{k-2} = \infty \right] \implies \left[ \lim_{i \rightarrow \infty} a(k)_i = \infty \right].$$

*Proof.* The case  $k = 2$  follows directly from the previous Lemma 3.13. Hence, assume  $k \geq 3$ . By Lemma 3.12 and the hypothesis we have that

$$\sum_{j=1}^{\infty} s_j \sum_{\sigma \in B_j^{k-2}} f(\sigma) = \infty.$$

Let  $N$  be constant from Lemma 3.13. Lemma 3.11 and the fact that  $0 \leq s_j - s_{N,j-1} \leq N+1$  imply that

$$\begin{aligned} 0 &\leq \sum_{j=1}^{\infty} s_j \sum_{\sigma \in B_j^{k-2}} f(\sigma) - \sum_{j=1}^{\infty} s_{N,j-1} \sum_{\sigma \in B_j^{k-2}} f(\sigma) \leq (N+1) \sum_{j=1}^{\infty} \sum_{\sigma \in B_j^{k-2}} f(\sigma) \\ &\leq (N+1) \sum_{j=1}^{\infty} \sum_{\sigma \in A_j^{k-2}} f(\sigma) < \infty. \end{aligned}$$

Hence we have that  $\sum_{j=1}^{\infty} s_{N,j-1} \sum_{\sigma \in B_j^{k-2}} f(\sigma) = \infty$ .

To complete the proof, we choose  $M > 0$ . Then there is  $i_0$  such that

$$\sum_{j=3}^{i_0+3-k} s_{N,j-1} \sum_{\sigma \in B_j^{k-2}} f(\sigma) > \frac{M+1}{C},$$

where  $C$  is the constants from Lemma 3.13. Now for each  $1 \leq j \leq i_0 + 3 - k$ , there is  $i_j$  such that

$$\sum_{\sigma \in B_{j,i_j}^{k-2}} f(\sigma) \geq \frac{-1}{C(i_0+3)(s_{i_0+3})} + \sum_{\sigma \in B_j^{k-2}} f(\sigma).$$

Now for all  $i \geq \max\{i_0, i_1, \dots, i_{i_0+3-k}\}$  we have that

$$\begin{aligned} a(k)_i &\geq C \sum_{j=3}^{i+3-k} s_{N,j-1} \sum_{\sigma \in B_{j,i}^{k-2}} f(\sigma) \\ &\geq C \sum_{j=3}^{i_0+3-k} s_{N,j-1} \sum_{\sigma \in B_{j,i_j}^{k-2}} f(\sigma) \\ &\geq C \sum_{j=3}^{i_0+3-k} s_{N,j-1} \left( \frac{-1}{C(i_0+3)(s_{i_0+3})} + \sum_{\sigma \in B_j^{k-2}} f(\sigma) \right) \\ &\geq -1 + C \sum_{j=3}^{i_0+3-k} s_{N,j-1} \sum_{\sigma \in B_j^{k-2}} f(\sigma) \\ &\geq -1 + C \frac{M+1}{C} = M, \end{aligned}$$

completing the proof.  $\square$

With these lemmas we can give a very good characterization of the probability space in the zero-one case. To state a compact theorem we introduce some elaborate notation to denote certain infinite graphs. Let  $T$  be a finite tree. Let  $F_T$  be the forest that consists of infinitely many copies of  $T$ , as components. Let  $F_n = \cup F_T$ , where the union is taken over all trees of size  $n$ . Note that  $F_1$  is the countably infinite independent set, and  $F_2$  is the countably infinite matching.

**Theorem 3.15.**

- i) If  $\sum_{i=1}^{\infty} \frac{d_i}{i} = \infty$ , then the space is concentrated, and the atom is a graph consisting copies of the  $\omega$ -tree, as many as the number of zeroes in the sequence.
- ii) Suppose  $\sum_{i=1}^{\infty} \frac{d_i}{i} < \infty$ . Let  $k = \min\{k \geq 2 : \sum_l d_l t_{l+1}^{k-2} < \infty\}$  (we allow  $k = \infty$ ). The space has infinitely many atoms, and all of them are of the form  $F \cup [\bigcup_{i < k} F_i]$  where  $F$  is a finite forest.

*Proof.* Part i) is an immediate consequence of Theorem 3.3 in the special case of zero-one sequences.

For part ii), let  $m < k$ ; Lemma 3.14 imply that  $a(m)_i \rightarrow \infty$ , that is the expected number of components of size  $m$  is infinity. But we will prove a stronger statement, namely that a.s. there are infinitely many components of size  $m$ . To see this, denote

by  $N(m)_i$  the number of components of size  $m$  spanned by  $\{v_0, \dots, v_i\}$ . Let  $E_i$  be the event that we create a component of size  $m$  at step  $i$ .

$$\begin{aligned} \Pr(E_i) &= \sum_{l=0}^{\infty} \Pr(E_i | N(m-1)_{i-1} = l) \Pr(N(m-1)_{i-1} = l) \\ &= \sum_{l=0}^{\infty} \frac{d_i(m-1)l}{i} \Pr(N(m-1)_{i-1} = l) = \frac{d_i(m-1)}{i} a(m-1)_{i-1} \end{aligned}$$

If we created a component of size  $m$  at step  $i$ , the probability that we don't ever destroy it is

$$q_i = \prod_{j=i+1}^{\infty} \left(1 - \frac{d_j m}{j}\right).$$

Note that  $q_i \rightarrow 1$  as  $i \rightarrow \infty$ .

Now let  $F_n$  denote the event that every component of size  $m$  that is created at or after the  $n$ th step is eventually destroyed. We will show that for all  $n$ ,  $\Pr(F_n) = 0$ . According to the calculations above

$$\Pr(F_n) = \prod_{i=n}^{\infty} \left(1 - \frac{d_i(m-1)}{i} a(m-1)_{i-1} q_i\right).$$

Since  $\liminf_{i \rightarrow \infty} (m-1)q_i > 0$

$$\sum \frac{d_j}{j} a(m-1)_{j-1} = \infty \implies \sum \frac{d_j(m-1)}{j} a(m-1)_{j-1} q_j = \infty$$

But by Lemma 3.9

$$\sum_{j=1}^i \frac{d_j}{j} a(m-1)_{j-1} \geq C a(m)_i \rightarrow \infty,$$

hence we conclude  $\Pr(F_n) = 0$ , which in turn implies that a.s. there are infinitely many components (which are all trees) of size  $m$ .

For any  $m' \geq k$ , Lemma 3.10 shows that the expected number of components of size  $m'$  is finite, therefore a.s. there are finitely many components of size  $m'$ .

There are two things remain to be proven to finish the proof of the theorem. First that if  $m < k$ , and  $T$  is a tree with  $|T| = m$ , then a.s. there are infinitely many components of the graph isomorphic to  $T$ . From the argument above, we have that a.s. there are infinitely many components of size  $m$ , and also, if  $C$  is a component,  $\Pr(C \cong T \mid |C| = m) > 0$ , so the statement follows.

The second thing is that a.s. there is no infinite component of the graph. From Theorem 3.3, we know that a.s. each vertex is of finite degree (i.e. the graph is locally finite). Every locally finite connected infinite graph contains a ray (see e.g. Proposition 8.2.1. in [1]). So Theorem 3.4 finishes the proof.  $\square$

### Corollary 3.16.

If  $\sum_{i=1}^{\infty} \frac{d_i s_i}{i} < \infty$ , then the space has infinitely many atoms, and all of them are of the form  $F \cup F_1 \cup F_2$  where  $F$  is a finite forest.

*Proof.*

$$\sum_{j=1}^{\infty} d_j t_{j+1} \leq \sum_{j=1}^{\infty} d_j t_j = \sum_{j=1}^{\infty} d_j \sum_{i=j}^{\infty} \frac{d_i}{i} = \sum_{i=1}^{\infty} \frac{d_i}{i} \sum_{j=1}^i d_j = \sum_{i=1}^{\infty} \frac{d_i s_i}{i} < \infty,$$

so Theorem 3.15 implies the statement.  $\square$

It is an interesting historical note that the authors proved this corollary well before the very powerful Theorem 3.15. Let us include that first, self-contained proof as well.

According to Lemma 3.5, every atom must have the property that beyond a vertex, each edge attaches back to vertices of current degree zero. That means that all, but finitely many vertices are in components  $K_1$  and  $K_2$  (singletons and arcs). The finitely many initial vertices will form a forest, so we have proven that any atom must be of the given form.

It remains to be shown that infinitely many of the graphs of the given form have positive probability. Consider any finite tree  $T$  on  $l$  vertices; we will show that with positive probability the graph will contain  $T$  as a component. Wait until we have just created the second component, then build  $T$  out of the second component with the next  $l$  vertices for which  $d_i = 1$ . (Waiting for two components is necessary to avoid the problem of having more than  $l$  1's in the beginning of the sequence.) This happens with some positive probability. Then apply Lemma 3.2 with  $V = V(T)$  and conclude that with positive probability  $T$  will be never hit again, so it will be a component of the graph.

The following theorem is a natural analogue of Corollary 2.6.

**Theorem 3.17.** *Fix  $0 < p < 1$ , and consider the double random process with  $d_i = 1$  with probability  $p$ , otherwise  $d_i = 0$  for  $i > 0$ . The process a.s. generates infinitely many copies of  $\omega$ -trees.*

*Proof.* We will prove that Theorem 3.15, i) is satisfied a.s. Let  $X_n = \sum_{i=0}^n d_i/i$ . On one hand  $\mu_n := E[X_n] \geq p \ln n$ . On the other hand,

$$\sigma_n^2 := \text{Var}[X_n] = \sum_{i=1}^n \text{Var} \left[ \frac{d_i}{i} \right] \leq \sum_{i=1}^n \frac{p - p^2}{i^2} \leq 2.$$

Fix  $M > 0$ . Using Chebyshev's inequality, if  $\mu_n > M$ ,

$$\Pr[X_n \leq M] \leq \Pr[|X_n - \mu_n| \geq \mu_n - M] \leq \frac{\sigma_n^2}{(\mu_n - M)^2} \leq \frac{2}{(p \ln n - M)^2} \rightarrow 0.$$

Hence, a.s.  $X_n \rightarrow \infty$ .  $\square$

## REFERENCES

- [1] Reinhard Diestel, *Graph theory*, 4 ed., Graduate Texts in Mathematics, vol. 173, Springer, 2010.
- [2] Paul Erdős and Alfréd Rényi, *Asymmetric graphs*, Acta Math. Acad. Sci. Hung. **14** (1963), 295–315.